CSE 547/Stat 548: Machine Learning for Big Data **Lecture**

Random Projections

Instructor: Sham Kakade

1 The Johnson-Lindenstrauss lemma

Theorem 1.1. *(Johnson-Lindenstrauss) Let $\epsilon \in (0, 1/2)$. Let $Q \subset \mathbb{R}^d$ be a set of n points and $k = \frac{20 \log n}{\epsilon^2}$. There exists a Lipshcitz mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$:*

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2$$

To prove JL, we appeal to the following lemma:

Theorem 1.2. *(Norm preservation) Let $x \in \mathbb{R}^d$. Assume that the entries in $A \subset \mathbb{R}^{k \times d}$ are sampled independently from $N(0, 1)$. Then,*

$$\Pr((1-\epsilon)\|x\|^2 \leq \|\frac{1}{\sqrt{k}}Ax\|^2 \leq (1+\epsilon)\|x\|^2) \geq 1 - 2e^{-(\epsilon^2-\epsilon^3)k/4}$$

The proof of the JL just appeals to the union bound:

Proof. The proof is constructive and is an example of the *probabilistic method*. Choose an f which is a *random projection*. Let $f = \frac{1}{\sqrt{k}}Ax$ where A is a $k \times d$ matrix, where each entry is sampled i.i.d from a Gaussian $N(0, 1)$.

Note there are $O(n^2)$ pairs of $u, v \in Q$. By the union bound,

$$\begin{aligned}
& \Pr(\exists u, v \text{ s.t. the following event fails: } (1-\epsilon)\|u-v\|^2 \leq \|\frac{1}{\sqrt{k}}A(u-v)\|^2 \leq (1+\epsilon)\|u-v\|^2) \\
\leq & \sum_{u,v \in Q} \Pr(\text{ s.t. the following event fails: } (1-\epsilon)\|u-v\|^2 \leq \|\frac{1}{\sqrt{k}}A(u-v)\|^2 \leq (1+\epsilon)\|u-v\|^2) \\
\leq & 2n^2 e^{-(\epsilon^2-\epsilon^3)k/4} \\
< & 1
\end{aligned}$$

the last step follows if we choose $k = \frac{20}{\epsilon^2}\log n)$.

Note that that the probability of finding a map f which satisfies the desired conditions is strictly greater than 0, so such a map must exist. (Aside: this proof technique is known as 'the probabilistic method' — note that the theorem is a deterministic statement while the proof is via a probabilistic argument.) □

Now let us prove the norm preservation lemma:

Proof. First let us show that for any $x \in \mathbb{R}^d$, we have that:

$$\mathbb{E}[\|\frac{1}{\sqrt{k}}Ax\|^2] = \mathbb{E}[\|x\|^2]\,.$$

To see this, let us examine the expected value of the entry $[Ax]_j^2$

$$\begin{aligned}
\mathbb{E}[[Ax]_j^2] &= \mathbb{E}[(\sum_{i=1}^{d} A_{i,j}x_i)^2] \\
&= \mathbb{E}[\sum_{i,i'} A_{i,j}A_{i',j}x_{i'}x_i] \\
&= \mathbb{E}[\sum_{i} A_{i,i}^2 x_i^2] \\
&= \sum_i x_i^2 \\
&= \|x\|^2
\end{aligned}$$

and note that:

$$\|\frac{1}{\sqrt{k}}Ax\|^2 = \frac{1}{k}\sum_{j=1}^{k}[Ax]_j^2$$

which proves the first claim (note that all we require for this proof is independence and unit variance in constructing A).

Note that above shows that $\tilde{Z}_j = [Ax]_j/\|x\|$ is distributed as $N(0,1)$, and $\tilde{Z}_j$ are independent. We now bound the failure probability of one side. By the union bound,

$$\begin{aligned}
\Pr(\|\frac{1}{\sqrt{k}}Ax\|^2 > (1+\epsilon)\|x\|^2) &= \Pr(\sum_{i=1}^{k}\tilde{Z}_i^2 > (1+\epsilon)k) \\
&= n^2 \Pr(\chi_k^2 > (1+\epsilon)k)
\end{aligned}$$

(where χ_k^2 is the chi-squared distribution with k degrees of freedom). Now we appeal to a concentration result below, which bounds this probability by:

$$\leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))$$

A similar argument handles the other side (and the factor of 2 in the bound). □

The following lemma for χ^2- distributions was used in the above proof.

Lemma 1.3. *We have that:*

$$\Pr(\chi_k^2 \geq (1+\epsilon)k) \leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))$$
$$\Pr(\chi_k^2 \leq (1-\epsilon)k) \leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))$$

Proof. Let $Z_1, Z_2, \ldots Z_k$ be i.i.d. $N(0,1)$ random variables. By Markov's inequality,

$$\begin{aligned}
\Pr(\chi_k^2 \geq (1+\epsilon)k) &= \Pr(\sum_{i=1}^{k} Z_i^2 > (1+\epsilon)k) \\
&= \Pr(e^{\lambda \sum_{i=1}^{k} Z_i^2} > e^{(1+\epsilon)k\lambda}) \\
&\leq \frac{\mathbb{E}[e^{\lambda \sum_{i=1}^{k} Z_i^2}]}{e^{(1+\epsilon)k\lambda}} \\
&= \frac{(\mathbb{E}[e^{\lambda Z_1^2}])^k}{e^{(1+\epsilon)k\lambda}} \\
&= e^{-(1+\epsilon)k\lambda} \left(\frac{1}{1-2\lambda}\right)^{k/2}
\end{aligned}$$

where the last step follows from evaluating the expectation, which holds for $0 < \lambda \leq 1/2$ (this expectation is just the moment generating function). Choosing $\lambda = \frac{\epsilon}{2(1+\epsilon)}$ which minimizes the above expression (and is less than $1/2$ as required), we have:

$$\begin{aligned}
\Pr(\chi_k^2 \geq (1+\epsilon)k) &= ((1+\epsilon)e^{-\epsilon})^{\frac{k}{2}}) \\
&\leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))
\end{aligned}$$

using the upper bound $1 + \epsilon \leq \exp(\epsilon - (\epsilon^2 - \epsilon^3)/2)$. The other bound is proved in a similar manner. □

The following lemma shows that nothing is fundamental about using Gaussian in particular. Many distributions with unit variance and certain boundedness properties (or higher order moment conditions) suffice.

Lemma 1.4. *Assume for $A \in \mathbb{R}^{\times} k \times d$ that each A_i, j is uniform on $\{-1, 1\}$. Then for any vector $x \in \mathbb{R}^d$:*

$$\Pr(\|\frac{1}{\sqrt{k}} Ax\|^2 \geq (1+\epsilon)\|x\|^2) \leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))$$
$$\Pr(\|\frac{1}{\sqrt{k}} Ax\|^2 \leq (1-\epsilon)\|x\|^2) \leq \exp(-\frac{k}{4}(\epsilon^2 - \epsilon^3))$$